

Large deviations for method-of-quantiles estimators of one-dimensional parameters*

Valeria Bignozzi[†]Claudio Macci[‡]Lea Petrella[§]

Abstract

We consider method-of-quantiles estimators of unknown parameters, namely the analogue of method-of-moments estimators obtained by matching empirical and theoretical quantiles at some probability level $\lambda \in (0, 1)$. The aim is to present large deviation results for these estimators as the sample size tends to infinity. We study in detail several examples; for specific models we discuss the choice of the optimal value of λ and we compare the convergence of the method-of-quantiles and method-of-moments estimators.

AMS Subject Classification: 60F10; 62F10; 62F12.

Keywords: location parameter; methods of moments; order statistics; scale parameter; skewness parameter.

1 Introduction

Estimation of parameters of statistical or econometric models is one of the main concerns in the parametric inference framework. When the probability law is specified (up to unknown parameters), the main tool to solve this problem is the Maximum Likelihood (ML) technique; on the other hand, whenever the assumption of a particular distribution is too restrictive, different solutions may be considered. For instance the Method of Moments (MM) and the Generalized Method of Moments (GMM) provide valuable alternative procedures; in fact the application of these methods only requires the knowledge of some moments.

A different approach is to consider the Method of Quantiles (MQ), that is the analogue of MM with quantiles; MQ estimators are obtained by matching the empirical percentiles with their theoretical counterparts at one or more probability levels. Inference via quantiles goes back to [1] where the authors consider an estimation problem for a three-parameter log-normal distribution; their approach consists in minimizing a suitable distance between the theoretical and empirical quantiles, see for instance [12]. Successive papers deal with the estimation of parameters of extreme value (see [8] and [11]), logistic (see [9]) and Weibull (see [10]) distributions. A more recent reference is [2] where several other distributions are studied. We also recall [6] where the authors consider an indirect inference method based on the simulation of theoretical quantiles, or a function of them,

*The support of Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica (INdAM) is acknowledged.

[†]Dipartimento di Statistica e Metodi Quantitativi, Università di Milano Bicocca, Via Bicocca degli Arcimboldi 8, I-20126 Milano, Italia. e-mail: valeria.bignozzi@unimib.it

[‡]Dipartimento di Matematica, Università di Roma Tor Vergata, Via della Ricerca Scientifica, I-00133 Roma, Italia. e-mail: macci@mat.uniroma2.it

[§]Dipartimento di Metodi e Modelli per l'Economia, il Territorio e la Finanza, Sapienza Università di Roma, Via del Castro Laurenziano 9, I-00161 Roma, Italia. e-mail: lea.petrella@uniroma1.it

when they are not available in a closed form. In [15], an iterative procedure based on ordinary least-squares estimation is proposed to compute MQ estimators; such estimators can be easily modified by adding a LASSO penalty term if a sparse representation is desired, or by restricting the matching within a given range of quantiles to match a part of the target distribution. Quantiles and empirical quantiles represent a key tool also in quantitative risk management, where they are studied under the name of Value-at-Risk (see for instance [13]).

In our opinion, MQ estimators deserve a deeper investigation because of several advantages. They allow to estimate parameters when the moments are not available and they are invariant with respect to increasing transformations; moreover they have less computational problems, and behave better when distributions are heavy-tailed or their supports vary with the parameters.

The aim of this paper is to present large deviation results for MQ estimators (as the sample size tends to infinity) for statistical models with one-dimensional unknown parameter $\theta \in \Theta$, where the parameter space Θ is a subset of the real line; thus we match empirical and theoretical quantiles at one probability level $\lambda \in (0, 1)$. The theory of large deviations is a collection of techniques which gives an asymptotic computation of small probabilities on an exponential scale (see e.g. [4] as a reference on this topic). Several examples of statistical models are considered throughout the paper, and some particular distributions are studied in detail. For most of the examples considered, we are able to find an explicit expression for the rate function which governs the large deviation principle of the MQ estimators and, when possible, our investigation provides the optimal λ that guarantees a faster convergence to the true parameter (see Definition 3.1). Further we compare MQ and MM estimators in terms of the local behavior of the rate functions around the true value of the parameter in the spirit of Remark 2.1. Which one of the estimators behaves better strictly depends on the type of parameter we have to estimate and varies upon distributions. However, we provide explicit examples (a part from the obvious ones where the MM estimators are not available) where MQ estimators are preferable.

We conclude with the outline of the paper. In Section 2 we recall some preliminaries. Sections 3 and 4 are devoted to the results for MQ and MM estimators, respectively. In Section 5 we present examples for different kind of parameters (e.g. scale, location, skewness, etc.), and for each example specific distributions are discussed in Section 6.

2 Preliminaries

In this section we present some preliminaries on large deviations and we provide a rigorous definition of the MQ estimators studied in this paper (see Definition 2.1 below).

2.1 Large deviations

We start with the concept of large deviation principle (LDP for short). A sequence of random variables $\{W_n : n \geq 1\}$ taking values on a topological space \mathcal{W} satisfies the LDP with rate function $I : \mathcal{W} \rightarrow [0, \infty]$ if I is a lower semi-continuous function,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P(W_n \in O) \geq - \inf_{w \in O} I(w) \text{ for all open sets } O$$

and

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P(W_n \in C) \leq - \inf_{w \in C} I(w) \text{ for all closed sets } C.$$

We also recall that a rate function I is said to be good if all its level sets $\{w \in \mathcal{W} : I(w) \leq \eta\} : \eta \geq 0\}$ are compact.

Remark 2.1 (Local comparison between rate functions around the unique common zero). *It is known that, if $I(w)$ uniquely vanish at some $w_0 \in \mathcal{W}$, then the sequence of random variables converges weakly to w_0 . Moreover, if we have two rate functions I_1 and I_2 which uniquely vanish at the same point $w_0 \in \mathcal{W}$, and if $I_1(w) > I_2(w) > 0$ in a neighborhood of w_0 (except w_0) then any sequence which satisfies the LDP with rate function I_1 converges to w_0 faster than any sequence which satisfies the LDP with rate function I_2 .*

We also recall a recent large deviation result on order statistics of i.i.d. random variables (see Proposition 2.1 below) which plays a crucial role in this paper. We start with the following condition.

Condition 2.1. *Let $\{X_n : n \geq 1\}$ be a sequence of i.i.d. real valued random variables with distribution function F , and assume that F is continuous and strictly increasing on (α, ω) , where $-\infty \leq \alpha < \omega \leq \infty$. Moreover let $\{k_n : n \geq 1\}$ be such that $k_n \in \{1, \dots, n\}$ for all $n \geq 1$ and $\lim_{n \rightarrow \infty} \frac{k_n}{n} = \lambda \in (0, 1)$.*

We introduce the following notation: for all $k \geq 1$, $X_{1:k} \leq \dots \leq X_{k:k}$ are the order statistics of the sample X_1, \dots, X_k ; for $p, q \in (0, 1)$ we set

$$H(p|q) := p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}, \quad (1)$$

that is the relative entropy of the Bernoulli distribution with parameter p with respect to the Bernoulli distribution with parameter q .

Proposition 2.1 (Theorem 3.2 in [7] for $\lambda \in (0, 1)$). *Assume that Condition 2.1 holds. Then $\{X_{k_n:n} : n \geq 1\}$ satisfies the LDP with good rate function $I_{\lambda,F}$ defined by*

$$I_{\lambda,F}(x) := \begin{cases} H(\lambda|F(x)) & \text{for } x \in (\alpha, \omega) \\ \infty & \text{otherwise.} \end{cases}$$

Remark 2.2 ($I''_{\lambda,F}(F^{-1}(\lambda))$ as the inverse of an asymptotic variance). *Theorem 7.1(c) in [3] states that, under suitable conditions, $\{\sqrt{n}(X_{k_n:n} - F^{-1}(\lambda)) : n \geq 1\}$ converges weakly to the centered Normal distribution with variance $\sigma^2 := \frac{\lambda(1-\lambda)}{(F'(F^{-1}(\lambda)))^2}$. Then, if we assume that F is twice differentiable, we can check that $I''_{\lambda,F}(F^{-1}(\lambda)) = \frac{1}{\sigma^2}$ with some computations.*

A more general formulation of Proposition 2.1 could be given also for $\lambda \in \{0, 1\}$ but, in view of the applications presented in this paper, we prefer to consider a restricted version of the result with $\lambda \in (0, 1)$ only. This restriction allows to have the goodness of the rate function $I_{\lambda,F}$ (see Remark 1 in [7]) which is needed to apply the contraction principle (see e.g. Theorem 4.2.1 in [4]).

2.2 MQ estimators

Here we present a rigorous definition of MQ estimators. In view of this, the next Condition 2.2 plays a crucial role.

Condition 2.2. *Let $\{F_\theta : \theta \in \Theta\}$ be a family of distribution functions where $\Theta \subset \mathbb{R}$ and, for all $\theta \in \Theta$, F_θ satisfies the same hypotheses of the distribution function F in Condition 2.1, for some $(\alpha_\theta, \omega_\theta)$. Moreover, for $\lambda \in (0, 1)$, consider the function $F_{(\bullet)}^{-1}(\lambda) : \Theta \rightarrow \mathcal{M}$, where $\mathcal{M} := \bigcup_{\theta \in \Theta} (\alpha_\theta, \omega_\theta)$, defined by*

$$[F_{(\bullet)}^{-1}(\lambda)](\theta) := F_\theta^{-1}(\lambda).$$

We assume that, for all $m \in \mathcal{M}$, the equation $[F_{(\bullet)}^{-1}(\lambda)](\theta) = m$ admits a unique solution (with respect to $\theta \in \Theta$) which will be denoted by $(F_{(\bullet)}^{-1}(\lambda))^{-1}(m)$.

Now we are ready to present the definition.

Definition 2.1. Assume that Condition 2.2 holds. Then $\left\{(F_{(\bullet)}^{-1}(\lambda))^{-1}(X_{[\lambda n]:n}) : n \geq 1\right\}$ is a sequence of MQ estimators (for the level $\lambda \in (0, 1)$).

Proposition 3.1 below provides the LDP for the sequence of estimators in Definition 2.1 (as the sample size n goes to infinity) when the true value of the parameter is $\theta_0 \in \Theta$. Actually we give a more general formulation in terms of $\left\{(F_{(\bullet)}^{-1}(\lambda))^{-1}(X_{k_n:n}) : n \geq 1\right\}$, where $\{k_n : n \geq 1\}$ is a sequence as in Condition 2.1.

3 Results for MQ estimators

In this section we prove the LDP for the sequence of estimators in Definition 2.1. Moreover we discuss some properties of the rate function; in particular Proposition 3.2 (combined with Remark 2.1 above) leads us to define a concept of optimal λ presented in Definition 3.1 below.

We start with our main result and, in view of this, we present the following notation:

$$h_{\lambda, \theta_0}(\theta) := F_{\theta_0}(F_{\theta}^{-1}(\lambda)) \text{ (for } F_{\theta}^{-1}(\lambda) \in (\alpha_{\theta_0}, \omega_{\theta_0})). \quad (2)$$

Proposition 3.1 (LD for MQ estimators). Assume that $\{k_n : n \geq 1\}$ is as in Condition 2.1 and that Condition 2.2 holds. Moreover assume that, for some $\theta_0 \in \Theta$, $\{X_n : n \geq 1\}$ are i.i.d. random variables with distribution function F_{θ_0} . Then, if the restriction of $(F_{(\bullet)}^{-1}(\lambda))^{-1}$ on $(\alpha_{\theta_0}, \omega_{\theta_0})$ is continuous, $\left\{(F_{(\bullet)}^{-1}(\lambda))^{-1}(X_{k_n:n}) : n \geq 1\right\}$ satisfies the LDP with good rate function I_{λ, θ_0} defined by

$$I_{\lambda, \theta_0}(\theta) := \begin{cases} \lambda \log \frac{\lambda}{h_{\lambda, \theta_0}(\theta)} + (1 - \lambda) \log \frac{1 - \lambda}{1 - h_{\lambda, \theta_0}(\theta)} & \text{for } \theta \in \Theta \text{ such that } F_{\theta}^{-1}(\lambda) \in (\alpha_{\theta_0}, \omega_{\theta_0}) \\ \infty & \text{otherwise,} \end{cases}$$

where $h_{\lambda, \theta_0}(\theta)$ is defined by (2).

Proof. Since the restriction of $(F_{(\bullet)}^{-1}(\lambda))^{-1}$ on $(\alpha_{\theta_0}, \omega_{\theta_0})$ is continuous, a straightforward application of the contraction principle yields the LDP of $\left\{(F_{(\bullet)}^{-1}(\lambda))^{-1}(X_{k_n:n}) : n \geq 1\right\}$ with good rate function I_{λ, θ_0} defined by

$$I_{\lambda, \theta_0}(\theta) := \inf \left\{ I_{\lambda, F_{\theta_0}}(x) : x \in (\alpha_{\theta_0}, \omega_{\theta_0}), (F_{(\bullet)}^{-1}(\lambda))^{-1}(x) = \theta \right\},$$

where $I_{\lambda, F_{\theta_0}}$ is the good rate function in Proposition 2.1, namely the good rate function defined by $I_{\lambda, F_{\theta_0}}(x) := H(\lambda | F_{\theta_0}(x))$, for $x \in (\alpha_{\theta_0}, \omega_{\theta_0})$. Moreover the set $\left\{x \in (\alpha_{\theta_0}, \omega_{\theta_0}) : (F_{(\bullet)}^{-1}(\lambda))^{-1}(x) = \theta\right\}$ has at most one element, namely

$$\left\{x \in (\alpha_{\theta_0}, \omega_{\theta_0}) : (F_{(\bullet)}^{-1}(\lambda))^{-1}(x) = \theta\right\} = \begin{cases} \{F_{\theta}^{-1}(\lambda)\} & \text{for } \theta \in \Theta \text{ such that } F_{\theta}^{-1}(\lambda) \in (\alpha_{\theta_0}, \omega_{\theta_0}) \\ \emptyset & \text{otherwise;} \end{cases}$$

thus we have $I_{\lambda, \theta_0}(\theta) = H(\lambda | F_{\theta_0}(F_{\theta}^{-1}(\lambda))) = H(\lambda | h_{\lambda, \theta_0}(\theta))$ for $\theta \in \Theta$ such that $F_{\theta}^{-1}(\lambda) \in (\alpha_{\theta_0}, \omega_{\theta_0})$, and $I_{\lambda, \theta_0}(\theta) = \infty$ otherwise. The proof is completed by taking into account the definition of the function H in (1). \square

Remark 3.1 (Rate function invariance with respect to increasing transformations). *Let $\{F_\theta : \theta \in \Theta\}$ be a family of distribution functions as in Condition 2.2 and assume that there exists an interval (α, ω) such that $(\alpha_\theta, \omega_\theta) = (\alpha, \omega)$ for all $\theta \in \Theta$. Moreover let $\psi : (\alpha, \omega) \rightarrow \mathbb{R}$ be a strictly increasing function. Then, if we consider the MQ estimators based on the sequence $\{\psi(X_n) : n \geq 1\}$ instead of $\{X_n : n \geq 1\}$, we can consider an adapted version of Proposition 3.1 with $(\psi(\alpha), \psi(\omega))$ in place of (α, ω) , $F_\theta \circ \psi^{-1}$ in place of F_θ and, as stated in Property 1.5.16 in [5], $\psi \circ F_\theta^{-1}$ in place of F_θ^{-1} . The LDP provided by this adapted version of Proposition 3.1 is governed by the rate function $I_{\lambda, \theta_0; \psi}$ defined by*

$$I_{\lambda, \theta_0; \psi}(\theta) := \begin{cases} H(\lambda | F_{\theta_0} \circ \psi^{-1}(\psi \circ F_\theta^{-1}(\lambda))) & \text{for } \theta \in \Theta \text{ such that } \psi \circ F_\theta^{-1}(\lambda) \in (\psi(\alpha), \psi(\omega)) \\ \infty & \text{otherwise} \end{cases}$$

instead of

$$I_{\lambda, \theta_0}(\theta) := \begin{cases} H(\lambda | F_{\theta_0}(F_\theta^{-1}(\lambda))) & \text{for } \theta \in \Theta \text{ such that } F_\theta^{-1}(\lambda) \in (\alpha, \omega) \\ \infty & \text{otherwise.} \end{cases}$$

One can easily realize that $I_{\lambda, \theta_0; \psi}$ and I_{λ, θ_0} coincide.

By taking into account the rate function in Proposition 3.1, it would be interesting to compare two rate functions I_{λ_1, θ_0} and I_{λ_2, θ_0} in the spirit of Remark 2.1 for a given pair $\lambda_1, \lambda_2 \in (0, 1)$; namely it would be interesting to have a strict inequality between I_{λ_1, θ_0} and I_{λ_2, θ_0} in a neighborhood of θ_0 (except θ_0).

Thus, if both rate functions are twice differentiable, I_{λ_1, θ_0} is locally larger (resp. smaller) than I_{λ_2, θ_0} around θ_0 if we have $I''_{\lambda_1, \theta_0}(\theta_0) > I''_{\lambda_2, \theta_0}(\theta_0)$ (resp. $I''_{\lambda_1, \theta_0}(\theta_0) < I''_{\lambda_2, \theta_0}(\theta_0)$). So it is natural to give an expression of $I''_{\lambda, \theta_0}(\theta_0)$ under suitable hypotheses.

Proposition 3.2 (An expression for $I''_{\lambda, \theta_0}(\theta_0)$). *Let I_{λ, θ_0} be the rate function in Proposition 3.1. Assume that $F_{\theta_0}(\cdot)$ and $F_{(\cdot)}^{-1}(\lambda)$ are twice differentiable. Then*

$$I''_{\lambda, \theta_0}(\theta_0) = \frac{(h'_{\lambda, \theta_0}(\theta_0))^2}{\lambda(1-\lambda)} = \frac{\{F'_{\theta_0}(F_{\theta_0}^{-1}(\lambda))\}^2}{\lambda(1-\lambda)} \left(\frac{d}{d\theta} F_\theta^{-1}(\lambda) \Big|_{\theta=\theta_0} \right)^2,$$

where $h_{\lambda, \theta_0}(\theta)$ is defined by (2).

Proof. One can easily check that

$$h_{\lambda, \theta_0}(\theta_0) = \lambda \text{ and } h'_{\lambda, \theta_0}(\theta_0) = F'_{\theta_0}(F_{\theta_0}^{-1}(\lambda)) \cdot \frac{d}{d\theta} F_\theta^{-1}(\lambda) \Big|_{\theta=\theta_0}.$$

Moreover after some computations we get

$$I'_{\lambda, \theta_0}(\theta) = h'_{\lambda, \theta_0}(\theta) \left(\frac{1-\lambda}{1-h_{\lambda, \theta_0}(\theta)} - \frac{\lambda}{h_{\lambda, \theta_0}(\theta)} \right)$$

and

$$I''_{\lambda, \theta_0}(\theta) = h''_{\lambda, \theta_0}(\theta) \left(\frac{1-\lambda}{1-h_{\lambda, \theta_0}(\theta)} - \frac{\lambda}{h_{\lambda, \theta_0}(\theta)} \right) + (h'_{\lambda, \theta_0}(\theta))^2 \left(\frac{\lambda}{h_{\lambda, \theta_0}^2(\theta)} + \frac{1-\lambda}{(1-h_{\lambda, \theta_0}(\theta))^2} \right).$$

Thus $I'_{\lambda, \theta_0}(\theta_0) = 0$ and $I''_{\lambda, \theta_0}(\theta_0) = (h'_{\lambda, \theta_0}(\theta_0))^2 \left(\frac{1}{\lambda} + \frac{1}{1-\lambda} \right) = \frac{(h'_{\lambda, \theta_0}(\theta_0))^2}{\lambda(1-\lambda)}$. The proof is completed by taking into account the expression of $h'_{\lambda, \theta_0}(\theta_0)$ above. \square

Finally, by taking into account Proposition 3.2 (and what we said before it), it is natural to consider the following

Definition 3.1. A value $\lambda_{\max} \in (0, 1)$ is said to be optimal if it maximizes $I''_{\lambda, \theta_0}(\theta_0)$, namely if we have $I''_{\lambda_{\max}, \theta_0}(\theta_0) = \sup_{\lambda \in (0, 1)} I''_{\lambda, \theta_0}(\theta_0)$.

4 Results for MM estimators

The aim of this section is to present a version of the above results for MM estimators; namely the LDP and an expression of $J''_{\theta_0}(\theta_0)$, where J_{θ_0} is the rate function which governs the LDP of MM estimators. In particular, when we compare MM and MQ estimators in terms of speed of convergence by referring to Remark 2.1, the value $J''_{\theta_0}(\theta_0)$ will be compared with $I''_{\lambda, \theta_0}(\theta_0)$ in Proposition 3.2.

We start with the following condition which allows us to define the MM estimators.

Condition 4.1. Let $\{F_\theta : \theta \in \Theta\}$ be a family of distribution functions as in Condition 2.2, and consider the function $\mu : \Theta \rightarrow \mathcal{M}$, where $\mathcal{M} := \bigcup_{\theta \in \Theta} (\alpha_\theta, \omega_\theta)$, defined by

$$\mu(\theta) := \int_{\alpha_\theta}^{\omega_\theta} x dF_\theta(x).$$

We assume that, for all $m \in \mathcal{M}$, the equation $\mu(\theta) = m$ admits a unique solution (with respect to $\theta \in \Theta$) which will be denoted by $\mu^{-1}(m)$.

From now on, in connection with this condition, we introduce the following function:

$$\Lambda_\theta^*(x) := \sup_{\gamma \in \mathbb{R}} \{\gamma x - \Lambda_\theta(\gamma)\}, \text{ where } \Lambda_\theta(\gamma) := \log \int_{\alpha_\theta}^{\omega_\theta} e^{\gamma x} dF_\theta(x). \quad (3)$$

It is well-known that, if $\{X_n : n \geq 1\}$ are i.i.d. random variables with distribution function F_θ , and if we set $\bar{X}_n := \frac{X_1 + \dots + X_n}{n}$ for all $n \geq 1$, then $\{\bar{X}_n : n \geq 1\}$ satisfies the LDP with rate function Λ_θ^* in (3) by Cramér Theorem on \mathbb{R} (see e.g. Theorem 2.2.3 in [4]).

Then we have the following result.

Proposition 4.1 (LD for MM estimators). Assume that Condition 4.1 holds. Moreover assume that, for some $\theta_0 \in \Theta$, $\{X_n : n \geq 1\}$ are i.i.d. random variables with distribution function F_{θ_0} .

(i) If $\mu^{-1}(m) := c_1 m + c_0$ for some $c_1, c_0 \in \mathbb{R}$ such that $c_1 \neq 0$, then $\{\mu^{-1}(\bar{X}_n) : n \geq 1\}$ satisfies the LDP with rate function J_{θ_0} defined by

$$J_{\theta_0}(\theta) := \begin{cases} \Lambda_{\theta_0}^*(\mu(\theta)) & \text{for } \theta \in \Theta \text{ such that } \mu(\theta) \in (\alpha_{\theta_0}, \omega_{\theta_0}) \\ \infty & \text{otherwise.} \end{cases}$$

(ii) If the restriction of μ^{-1} on $(\alpha_{\theta_0}, \omega_{\theta_0})$ is continuous and if $\Lambda_{\theta_0}^*$ is a good rate function, the same LDP holds and J_{θ_0} is a good rate function.

Proof. (i) In this case $\mu(\theta) := \frac{\theta - c_0}{c_1}$ and $\{\mu^{-1}(\bar{X}_n) : n \geq 1\}$ is again a sequence of empirical means of i.i.d. random variables. Then the LDP still holds by Cramér Theorem on \mathbb{R} , and the rate function J_{θ_0} is defined by

$$J_{\theta_0}(\theta) := \sup_{\gamma \in \mathbb{R}} \{\gamma \theta - \Lambda_{\theta_0}(c_1 \gamma) - \gamma c_0\},$$

which yields

$$J_{\theta_0}(\theta) = \sup_{\gamma \in \mathbb{R}} \left\{ c_1 \gamma \frac{\theta - c_0}{c_1} - \Lambda_{\theta_0}(c_1 \gamma) \right\} = \sup_{\gamma \in \mathbb{R}} \{ c_1 \gamma \mu(\theta) - \Lambda_{\theta_0}(c_1 \gamma) \} = \Lambda_{\theta_0}^*(\mu(\theta)),$$

as desired.

(ii) Since the restriction of the function μ^{-1} on $(\alpha_{\theta_0}, \omega_{\theta_0})$ is continuous and $\Lambda_{\theta_0}^*$ is a good rate function, a straightforward application of the contraction principle yields the LDP of $\{\mu^{-1}(\bar{X}_n) : n \geq 1\}$ with good rate function J_{θ_0} defined by

$$J_{\theta_0}(\theta) := \inf \{ \Lambda_{\theta_0}^*(x) : x \in (\alpha_{\theta_0}, \omega_{\theta_0}), \mu^{-1}(x) = \theta \}.$$

Moreover the set $\{x \in (\alpha_{\theta_0}, \omega_{\theta_0}) : \mu^{-1}(x) = \theta\}$ has at most one element, namely

$$\{x \in (\alpha_{\theta_0}, \omega_{\theta_0}) : \mu^{-1}(x) = \theta\} = \begin{cases} \{\mu(\theta)\} & \text{for } \theta \in \Theta \text{ such that } \mu(\theta) \in (\alpha_{\theta_0}, \omega_{\theta_0}) \\ \emptyset & \text{otherwise;} \end{cases}$$

thus we have $J_{\theta_0}(\theta) = \Lambda_{\theta_0}^*(\mu(\theta))$ for $\theta \in \Theta$ such that $\mu(\theta) \in (\alpha_{\theta_0}, \omega_{\theta_0})$, and $J_{\theta_0}(\theta) = \infty$ otherwise. \square

Now, in the spirit of Remark 2.1, it would be interesting to have a local strict inequality between the rate function $I_{\lambda, \theta_0}(\theta)$ in Proposition 3.1 for MQ estimators (for some $\lambda \in (0, 1)$), and the rate function $J_{\theta_0}(\theta)$ in Proposition 4.1 for MM estimators.

Then we can repeat the same arguments which led us to Proposition 3.2. Namely, if both rate functions J_{θ_0} and I_{λ, θ_0} (for some $\lambda \in (0, 1)$) are twice differentiable, J_{θ_0} is locally larger (resp. smaller) than I_{λ, θ_0} around θ_0 if $J_{\theta_0}''(\theta_0) > I_{\lambda, \theta_0}''(\theta_0)$ (resp. $J_{\theta_0}''(\theta_0) < I_{\lambda, \theta_0}''(\theta_0)$). So it is natural to give an expression of $J_{\theta_0}''(\theta_0)$ under suitable hypotheses.

Proposition 4.2 (An expression for $J_{\theta_0}''(\theta_0)$). *Let J_{θ_0} be the rate function in Proposition 4.1. Assume that, for all $\theta \in \Theta$, the function Λ_θ in (3) is finite in a neighborhood of the origin $\gamma = 0$ and that $\mu(\cdot)$ is twice differentiable. Then $J_{\theta_0}''(\theta_0) = \frac{(\mu'(\theta_0))^2}{\sigma^2(\theta_0)}$, where $\sigma^2(\cdot)$ is the variance function.*

Proof. One can easily check that

$$J'_{\theta_0}(\theta) = (\Lambda_{\theta_0}^*)'(\mu(\theta))\mu'(\theta) \text{ and } J''_{\theta_0}(\theta) = (\Lambda_{\theta_0}^*)''(\mu(\theta))(\mu'(\theta))^2 + \mu''(\theta)(\Lambda_{\theta_0}^*)'(\mu(\theta)).$$

Then we can conclude noting that $(\Lambda_{\theta_0}^*)''(\mu(\theta)) = \frac{1}{\sigma^2(\theta)}$ and $(\Lambda_{\theta_0}^*)'(\mu(\theta)) = 0$. \square

Remark 4.1 (On the functions Λ_θ and Λ_θ^* in (3)). *The function Λ_θ is finite in a neighborhood of the origin $\gamma = 0$ when we deal with empirical means (of i.i.d. random variables) with light-tailed distribution. Typically Λ_θ^* is a good rate function only in this case.*

5 Examples

The aim of this section is to present several examples of statistical models with unknown parameter $\theta \in \Theta$, where $\Theta \subset \mathbb{R}$; in all the examples we always deal with one-dimensional parameters assuming all the others to be known.

Let us briefly introduce the examples presented below. We investigate distributions with scale parameter in Example 1, with location parameter in Example 2, and with skewness parameter in Example 3. We remark that in Example 3 we use the epsilon-Skew-Normal distribution defined in [14]; this choice is motivated by the availability of an explicit expression of the inverse of the

distribution function giving us the possibility of obtaining explicit formulas. Moreover we present Example 4 with Pareto distributions, which allows to give a concrete illustration of the content of Remark 3.1. In all these statistical models the intervals $\{(\alpha_\theta, \omega_\theta) : \theta \in \Theta\}$ do not depend on θ and we simply write (α, ω) . Finally we present Example 5 where we have $(\alpha_\theta, \omega_\theta) = (0, \theta)$ for $\theta \in \Theta := (0, \infty)$; namely for this example θ is a right-endpoint parameter.

In all examples (except Example 4) we give a formula for $I''_{\lambda, \theta_0}(\theta_0)$ (as a consequence of Proposition 3.2) which will be used for the local comparisons between rate functions (in the spirit of Remark 2.1) analyzed in Section 6.

In what follows we say that a distribution function F on \mathbb{R} has the *symmetry property* if it is a distribution function of a symmetric random variable, i.e. if $F(x) = 1 - F(-x)$ for all $x \in \mathbb{R}$. In such a case we have $F^{-1}(\lambda) = -F^{-1}(1 - \lambda)$ for all $\lambda \in (0, 1)$.

Example 1 (Statistical model with a scale parameter $\theta \in \Theta := (0, \infty)$). Let F_θ be defined by

$$F_\theta(x) := G\left(\frac{x}{\theta}\right) \text{ for } x \in (\alpha, \omega),$$

where G is a strictly increasing distribution function on $(\alpha, \omega) = (0, \infty)$ or $(\alpha, \omega) = (-\infty, \infty)$. Then

$$F_\theta^{-1}(\lambda) := \theta G^{-1}(\lambda) \text{ and } h_{\lambda, \theta_0}(\theta) = F_{\theta_0}(F_\theta^{-1}(\lambda)) = G\left(\frac{\theta}{\theta_0} \cdot G^{-1}(\lambda)\right);$$

it is important to remark that, when $(\alpha, \omega) = (-\infty, \infty)$, the value $\lambda = G(0)$ (which yields $G^{-1}(\lambda) = 0$) is not allowed. Now we give a list of some specific examples studied in this paper.

For the case $(\alpha, \omega) = (0, \infty)$ we consider the Weibull distribution:

$$G(x) := 1 - \exp(-x^\rho) \text{ (where } \rho > 0) \text{ and } G^{-1}(\lambda) := (-\log(1 - \lambda))^{1/\rho}. \quad (4)$$

We also give some specific examples where $(\alpha, \omega) = (-\infty, \infty)$ and, in each case, $\eta \in \mathbb{R}$ is a known location parameter (and the not-allowed value $\lambda = G(0)$ depends on η): the Normal distribution

$$G(x) := \Phi(x - \eta) \text{ and } G^{-1}(\lambda) := \eta + \Phi^{-1}(\lambda), \quad (5)$$

where Φ is the standard Normal distribution function; the Cauchy distribution

$$G(x) := \frac{1}{\pi} \left(\arctan(x - \eta) + \frac{\pi}{2} \right) \text{ and } G^{-1}(\lambda) := \eta + \tan\left(\left(\lambda - \frac{1}{2}\right)\pi\right); \quad (6)$$

the logistic distribution

$$G(x) := \frac{1}{1 + e^{-(x-\eta)}} \text{ and } G^{-1}(\lambda) := \eta - \log\left(\frac{1}{\lambda} - 1\right); \quad (7)$$

the Gumbel distribution

$$G(x) := \exp(-e^{-(x-\eta)}) \text{ and } G^{-1}(\lambda) := \eta - \log(-\log \lambda). \quad (8)$$

If G is twice differentiable we have

$$I''_{\lambda, \theta_0}(\theta_0) = \frac{\{G'(G^{-1}(\lambda))G^{-1}(\lambda)\}^2}{\lambda(1 - \lambda)\theta_0^2} \quad (9)$$

by Proposition 3.2; so, if it is possible to find an optimal λ_{\max} , such a value does not depend on θ_0 (on the contrary it could depend on the known location parameter η as we shall see in Section 6). Moreover one can check that $I''_{\lambda, \theta_0}(\theta_0) = 0$ if we consider the not-allowed value $\lambda = G(0) \in (0, 1)$ (when $(\alpha, \omega) = (-\infty, \infty)$) because $G^{-1}(\lambda) = 0$, and that $I''_{\lambda, \theta_0}(\theta_0) = I''_{1-\lambda, \theta_0}(\theta_0)$ (for all $\lambda \in (0, 1)$) if G is symmetric as it happens, for instance, in (5), (6) and (7) with $\eta = 0$.

Example 2 (Statistical model with a location parameter $\theta \in \Theta := (-\infty, \infty)$). Let F_θ be defined by

$$F_\theta(x) := G(x - \theta) \text{ for } x \in (\alpha, \omega) = (-\infty, \infty),$$

where G is a strictly increasing distribution function on $(\alpha, \omega) = (-\infty, \infty)$. Then

$$F_\theta^{-1}(\lambda) := \theta + G^{-1}(\lambda) \text{ and } h_{\lambda, \theta_0}(\theta) = F_{\theta_0}(F_\theta^{-1}(\lambda)) = G(\theta + G^{-1}(\lambda) - \theta_0).$$

We give some specific examples studied in this paper and, in each case, $s > 0$ is a known scale parameter: the Normal distribution

$$G(x) := \Phi\left(\frac{x}{s}\right) \text{ and } G^{-1}(\lambda) := s \cdot \Phi^{-1}(\lambda); \quad (10)$$

the Cauchy distribution

$$G(x) := \frac{1}{\pi} \left(\arctan \frac{x}{s} + \frac{\pi}{2} \right) \text{ and } G^{-1}(\lambda) := s \cdot \tan \left(\left(\lambda - \frac{1}{2} \right) \pi \right); \quad (11)$$

the logistic distribution

$$G(x) := \frac{1}{1 + e^{-x/s}} \text{ and } G^{-1}(\lambda) := -s \cdot \log \left(\frac{1}{\lambda} - 1 \right); \quad (12)$$

the Gumbel distribution

$$G(x) := \exp(-e^{-x/s}) \text{ and } G^{-1}(\lambda) := -s \cdot \log(-\log \lambda). \quad (13)$$

If G is twice differentiable we have

$$I''_{\lambda, \theta_0}(\theta_0) = \frac{\{G'(G^{-1}(\lambda))\}^2}{\lambda(1 - \lambda)} \quad (14)$$

by Proposition 3.2; so, if it is possible to find an optimal λ_{\max} , such a value does not depend on θ_0 and on the known scale parameter s . Moreover one can check that $I''_{\lambda, \theta_0}(\theta_0) = I''_{1-\lambda, \theta_0}(\theta_0)$ (for all $\lambda \in (0, 1)$) if G has the symmetry property (as happens for G in (10), (11) and (12), and not for G in (13)).

Example 3 (Statistical model with a skewness parameter $\theta \in \Theta := (-1, 1)$). Let F_θ be defined by

$$F_\theta(x) := \begin{cases} (1 + \theta)G(\frac{x}{1+\theta}) & \text{for } x \leq 0 \\ \theta + (1 - \theta)G(\frac{x}{1-\theta}) & \text{for } x > 0, \end{cases} \text{ with } x \in (\alpha, \omega) = (-\infty, \infty),$$

where G is a strictly increasing distribution function on $(\alpha, \omega) = (-\infty, \infty)$ with the symmetry property. Then

$$F_\theta^{-1}(\lambda) := \begin{cases} (1 + \theta)G^{-1}(\frac{\lambda}{1+\theta}) & \text{for } \lambda \in (0, \frac{1+\theta}{2}] \\ (1 - \theta)G^{-1}(\frac{\lambda-\theta}{1-\theta}) & \text{for } \lambda \in (\frac{1+\theta}{2}, 1) \end{cases}$$

and

$$h_{\lambda, \theta_0}(\theta) = F_{\theta_0}(F_\theta^{-1}(\lambda)) = \begin{cases} (1 + \theta_0)G\left(\frac{1+\theta}{1+\theta_0}G^{-1}(\frac{\lambda}{1+\theta})\right) & \text{for } \theta \geq 2\lambda - 1 \\ \theta_0 + (1 - \theta_0)G\left(\frac{1-\theta}{1-\theta_0}G^{-1}(\frac{\lambda-\theta}{1-\theta})\right) & \text{for } \theta < 2\lambda - 1. \end{cases}$$

We can consider the same specific examples presented in Example 2, i.e. the functions G in (10), (11) and (12) for some known scale parameter $s > 0$.

If G is twice differentiable and $G''(0) = 0$ we have

$$I''_{\lambda, \theta_0}(\theta_0) = \begin{cases} \frac{1}{\lambda(1-\lambda)} \left[G'(G^{-1}(\frac{\lambda}{1+\theta_0})) \left(G^{-1}(\frac{\lambda}{1+\theta_0}) - \frac{\lambda}{1+\theta_0} (G^{-1})'(\frac{\lambda}{1+\theta_0}) \right) \right]^2 & \text{for } \lambda \in (0, \frac{1+\theta_0}{2}] \\ \frac{1}{\lambda(1-\lambda)} \left[G'(G^{-1}(\frac{\lambda-\theta_0}{1-\theta_0})) \left(-G^{-1}(\frac{\lambda-\theta_0}{1-\theta_0}) + \frac{\lambda-1}{1-\theta_0} (G^{-1})'(\frac{\lambda-\theta_0}{1-\theta_0}) \right) \right]^2 & \text{for } \lambda \in (\frac{1+\theta_0}{2}, 1) \end{cases}$$

by Proposition 3.2, and therefore

$$I''_{\lambda, \theta_0}(\theta_0) = \begin{cases} \frac{1}{\lambda(1-\lambda)} \left[G'(G^{-1}(\frac{\lambda}{1+\theta_0})) G^{-1}(\frac{\lambda}{1+\theta_0}) - \frac{\lambda}{1+\theta_0} \right]^2 & \text{for } \lambda \in (0, \frac{1+\theta_0}{2}] \\ \frac{1}{\lambda(1-\lambda)} \left[-G'(G^{-1}(\frac{\lambda-\theta_0}{1-\theta_0})) G^{-1}(\frac{\lambda-\theta_0}{1-\theta_0}) + \frac{\lambda-1}{1-\theta_0} \right]^2 & \text{for } \lambda \in (\frac{1+\theta_0}{2}, 1); \end{cases} \quad (15)$$

so one can expect that, if it is possible to find an optimal λ_{\max} , such a value depends on θ_0 (this is what happens in Section 6). Moreover one can check that $I''_{\lambda, \theta_0}(\theta_0)$ in (15) does not depend on s , $I''_{1/2, 0}(0) = 1$ and $I''_{\lambda, \theta_0}(\theta_0) = I''_{1-\lambda, -\theta_0}(-\theta_0)$ (for all $\lambda \in (0, 1)$).

Example 4 (Statistical model with Pareto distributions with $\theta \in \Theta := (0, \infty)$). Let F_θ be defined by

$$F_\theta(x) := 1 - x^{-1/\theta} \text{ for } x \in (\alpha, \omega) = (1, \infty).$$

Then

$$F_\theta^{-1}(\lambda) := e^{-\theta \log(1-\lambda)} = (1-\lambda)^{-\theta} \text{ and } h_{\lambda, \theta_0}(\theta) = F_{\theta_0}(F_\theta^{-1}(\lambda)) = 1 - (1-\lambda)^{\theta/\theta_0}. \quad (16)$$

We remark that, if we consider Example 1 with G as in (4) with $\rho = 1$, namely

$$\tilde{F}_\theta(x) = 1 - e^{-x} \text{ for } (\tilde{\alpha}, \tilde{\omega}) = (0, \infty),$$

we can refer to Remark 3.1 with

$$\psi(x) := e^x \text{ for } x \in (\tilde{\alpha}, \tilde{\omega}) := (0, \infty)$$

(note that $(\psi(\tilde{\alpha}), \psi(\tilde{\omega})) = (1, \infty) = (\alpha, \omega)$). Then, as pointed out in Remark 3.1, $I_{\lambda, \theta_0; \psi}$ and I_{λ, θ_0} coincide; in fact $h_{\lambda, \theta_0}(\theta)$ in (16) meets the analogue expression in Example 1 with G as in (4) with $\rho = 1$, namely

$$\tilde{F}_{\theta_0}(\tilde{F}_\theta^{-1}(\lambda)) = G\left(\frac{\theta}{\theta_0} \cdot G^{-1}(\lambda)\right) = 1 - \exp\left(-\frac{\theta}{\theta_0} \cdot (-\log(1-\lambda))\right) = 1 - (1-\lambda)^{\theta/\theta_0}.$$

Example 5 (Statistical model with a “right endpoint” parameter $\theta \in \Theta := (0, \infty)$). Let F_θ be defined by

$$F_\theta(x) := \frac{G(x)}{G(\theta)} \text{ for } x \in (\alpha_\theta, \omega_\theta) := (0, \theta),$$

where $G : [0, \infty) \rightarrow [0, \infty)$ is a strictly increasing function such that $G(0) = 0$. Then

$$F_\theta^{-1}(\lambda) := G^{-1}(\lambda G(\theta)) \text{ and } h_{\lambda, \theta_0}(\theta) = F_{\theta_0}(F_\theta^{-1}(\lambda)) = \frac{\lambda G(\theta)}{G(\theta_0)} \text{ (for } \lambda G(\theta) \in (0, G(\theta_0))).$$

Moreover, after some computations, we get

$$I_{\lambda, \theta_0}(\theta) := \begin{cases} \lambda \log \frac{G(\theta_0)}{G(\theta)} + (1-\lambda) \log \frac{(1-\lambda)G(\theta_0)}{G(\theta_0) - \lambda G(\theta)} & \text{for } 0 < \theta < G^{-1}\left(\frac{G(\theta_0)}{\lambda}\right) \\ \infty & \text{otherwise.} \end{cases}$$

As a specific example we can consider $G(x) = x$ (for all x); in such a case F_θ is the distribution function concerning the uniform distribution on $(0, \theta)$. Finally, if G is twice differentiable, we have

$$I''_{\lambda, \theta_0}(\theta_0) = \frac{\lambda(G'(\theta_0))^2}{(1-\lambda)G^2(\theta_0)} \quad (17)$$

by Proposition 3.2.

6 Local comparisons between rate functions for some examples

In this section we analyze the examples presented in Section 5. We consider local comparisons between rate functions in the spirit of Remark 2.1 and, more precisely, the following two issues.

- A discussion on the choice of optimal values of λ in order to get the best rate of convergence. More precisely we study the behavior of $I''_{\lambda, \theta_0}(\theta_0)$ (varying $\lambda \in (0, 1)$) in order to find an optimal λ_{\max} in the sense of Definition 3.1.
- The comparison of the convergence of the MQ estimators and of the MM estimators. More precisely, when we deal with MM estimators, we compare $I''_{\lambda, \theta_0}(\theta_0)$ in Proposition 3.2 (for some $\lambda \in (0, 1)$) and $J''_{\theta_0}(\theta_0)$ in Proposition 4.2; obviously, when we have an optimal λ_{\max} , we take $\lambda = \lambda_{\max}$. In the single case presented below where MM estimators are not defined, we compare the convergence of the MQ estimators and of suitable GMM estimators.

We find at least an optimal λ_{\max} for all examples except for Example 5 (where we should consider $\lambda = 1$). For several particular examples where $(\alpha, \omega) = (-\infty, \infty)$ we have $I''_{\lambda, \theta_0}(\theta_0) = I''_{1-\lambda, \theta_0}(\theta_0)$ for all $\lambda \in (0, 1)$; then, in those cases, we have the optimal value $\lambda_{\max} = 1/2$ or, by symmetry, two distinct optimal values $\lambda_{\max,1}$ and $\lambda_{\max,2} = 1 - \lambda_{\max,1}$.

In view of what follows it is useful to consider two suitable values $\tilde{\lambda}_1, \tilde{\lambda}_2 \in (0, 1)$ presented in the next Lemma 6.1. The value $\tilde{\lambda}_1$ appears in the computations for the Weibull distribution in Example 1 (and also in the computations for Example 4 as a trivial consequence), while the $\tilde{\lambda}_2$ appears in the computations for the Gumbel distribution in both Examples 1 and 2; however, interestingly, Lemma 6.1(iii) shows the close relationship between $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$.

Lemma 6.1 (The values $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$). *The following statements hold.*

- (i) Let f_1 be the function defined by $f_1(\lambda) := \frac{(1-\lambda)(\log(1-\lambda))^2}{\lambda}$. Then $\sup_{\lambda \in (0,1)} f_1(\lambda) = f_1(\tilde{\lambda}_1)$, where $\tilde{\lambda}_1 \simeq 0.7968$ is the unique value $(1/2, 1)$ such that $-2\tilde{\lambda}_1 - \log(1 - \tilde{\lambda}_1) = 0$.
- (ii) Let f_2 be the function defined by $f_2(\lambda) := \frac{\lambda(\log \lambda)^2}{1-\lambda}$. Then $\sup_{\lambda \in (0,1)} f_2(\lambda) = f_2(\tilde{\lambda}_2)$, where $\tilde{\lambda}_2 \simeq 0.2032$ is the unique value $(0, 1/2)$ such that $\log \tilde{\lambda}_2 + 2 - 2\tilde{\lambda}_2 = 0$.
- (iii) We have $\tilde{\lambda}_1 + \tilde{\lambda}_2 = 1$.

Proof. (i) The derivative of f_1 is

$$\frac{d}{d\lambda} f_1(\lambda) = \frac{\log(1-\lambda)[-2\lambda - \log(1-\lambda)]}{\lambda^2}$$

and $-2\lambda - \log(1-\lambda) < 0$ (resp. $-2\lambda - \log(1-\lambda) > 0$) if $\lambda < \tilde{\lambda}_1$ (resp. $\lambda > \tilde{\lambda}_1$); therefore we have $\frac{d}{d\lambda} f_1(\lambda) > 0$ for $\lambda \in (0, \tilde{\lambda}_1)$ and $\frac{d}{d\lambda} f_1(\lambda) < 0$ for $\lambda \in (\tilde{\lambda}_1, 1)$.

(ii) The derivative of f_2 is

$$\frac{d}{d\lambda} f_2(\lambda) = \frac{\log \lambda \cdot [\log \lambda + 2 - 2\lambda]}{(1-\lambda)^2}$$

and $\log \lambda + 2 - 2\lambda < 0$ (resp. $\log \lambda + 2 - 2\lambda > 0$) if $\lambda < \tilde{\lambda}_2$ (resp. $\lambda > \tilde{\lambda}_2$); therefore we have $\frac{d}{d\lambda} f_2(\lambda) > 0$ for $\lambda \in (0, \tilde{\lambda}_2)$ and $\frac{d}{d\lambda} f_2(\lambda) < 0$ for $\lambda \in (\tilde{\lambda}_2, 1)$.

(iii) The equality $\tilde{\lambda}_1 + \tilde{\lambda}_2 = 1$ can be easily checked by taking into account how the values $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ are defined in statements (i) and (ii). \square

6.1 Example 1

In this section we consider the particular example of the Weibull distribution with $(\alpha, \omega) = (0, \infty)$, and the particular examples with $(\alpha, \omega) = (-\infty, \infty)$. In each part we analyze the MQ estimators, and we conclude with the MM estimators.

Analysis of MQ estimators for Weibull distribution. Here we consider G in (4). By (9) we have

$$I''_{\lambda, \theta_0}(\theta_0) = \frac{\rho^2(1-\lambda)(\log(1-\lambda))^2}{\lambda\theta_0^2}.$$

Then we have a unique optimal value $\lambda_{\max} = \tilde{\lambda}_1$ (for every ρ and θ_0), where $\tilde{\lambda}_1$ is the value in Lemma 6.1(i); in fact, if we consider the function f_1 in that lemma, we have $I''_{\lambda, \theta_0}(\theta_0) = \frac{\rho^2 f_1(\lambda)}{\theta_0^2}$.

MM versus MQ estimators for Weibull distribution. We start with the analysis of MM estimators. We have $\mu(\theta) := \theta\Gamma(1+1/\rho)$, and therefore $\mu^{-1}(m) := \frac{m}{\Gamma(1+1/\rho)}$, where Γ is the usual Gamma function. Thus, by Proposition 4.1(i), $\{\mu^{-1}(\bar{X}_n) : n \geq 1\}$ satisfies the LDP with rate function J_{θ_0} with $c_1 = \frac{1}{\Gamma(1+1/\rho)}$ and $c_0 = 0$. In what follows we consider the light-tailed case $\rho \geq 1$ and the heavy-tailed case $\rho \in (0, 1)$; some more details on the exponential distribution case $\rho = 1$ are given in Remark 6.1.

Light-tailed case (namely $\rho \geq 1$). The rate functions J_{θ_0} and $\Lambda_{\theta_0}^*$ are good and we can refer to the comparison between

$$J''_{\theta_0}(\theta_0) = \frac{\Gamma^2(1+1/\rho)}{\theta_0^2 [\Gamma(1+2/\rho) - \Gamma^2(1+1/\rho)]}$$

(this value is a consequence of Proposition 4.2 noting that $\sigma^2(\theta) = \theta^2 [\Gamma(1+2/\rho) - \Gamma^2(1+1/\rho)]$) and, for the optimal value $\lambda_{\max} = \tilde{\lambda}_1$ in Lemma 6.1(i),

$$I''_{\tilde{\lambda}_1, \theta_0}(\theta_0) = \frac{(1-\tilde{\lambda}_1)(\log(1-\tilde{\lambda}_1))^2}{\tilde{\lambda}_1\theta_0^2} = \frac{(1-\tilde{\lambda}_1)(2\tilde{\lambda}_1)^2}{\tilde{\lambda}_1\theta_0^2} = \frac{4\tilde{\lambda}_1(1-\tilde{\lambda}_1)}{\theta_0^2}.$$

We remark that $J''_{\theta_0}(\theta_0)$ is an increasing function of $\rho \in (0, \infty)$; in fact, if we set

$$a(\rho) := \Gamma'(1+2/\rho)\Gamma(1+1/\rho) - \Gamma(1+2/\rho)\Gamma'(1+1/\rho),$$

we have $a(\rho) > 0$ (noting that $\frac{\Gamma'(1+2/\rho)}{\Gamma(1+2/\rho)} > \frac{\Gamma'(1+1/\rho)}{\Gamma(1+1/\rho)}$ because the digamma function $x \mapsto \frac{\Gamma'(x)}{\Gamma(x)}$ is increasing on $(0, \infty)$) and therefore

$$\frac{d}{d\rho} \left(\frac{\Gamma^2(1+1/\rho)}{\theta_0^2 [\Gamma(1+2/\rho) - \Gamma^2(1+1/\rho)]} \right) = \frac{2\Gamma(1+1/\rho)a(\rho)}{\rho^2\theta_0^2 \left(\frac{\Gamma(1+2/\rho)}{\Gamma^2(1+1/\rho)} - 1 \right)^2 \Gamma^4(1+1/\rho)} > 0.$$

Thus, for all $\rho \geq 1$, MM estimators converge faster than every MQ estimators because

$$J''_{\theta_0}(\theta_0) \geq \frac{1}{\theta_0^2} > \frac{4\tilde{\lambda}_1(1-\tilde{\lambda}_1)}{\theta_0^2} = I''_{\tilde{\lambda}_1, \theta_0}(\theta_0)$$

noting that $\inf_{\rho \geq 1} J''_{\theta_0}(\theta_0) = J''_{\theta_0}(\theta_0)|_{\rho=1} = \frac{1}{\theta_0^2}$ and $4\tilde{\lambda}_1(1-\tilde{\lambda}_1) \simeq 0.6476$.

Heavy-tailed case (namely $\rho \in (0, 1)$). The rate functions J_{θ_0} and $\Lambda_{\theta_0}^*$ are not good (we recall Remark 4.1 presented above). We can say that $J_{\theta_0}(\theta) = 0$ for $\theta \geq \theta_0$; thus $I_{\lambda, \theta_0}(\theta) > J_{\theta_0}(\theta)$ for $\theta > \theta_0$ (for all $\lambda \in (0, 1)$). Then we have to compare $I_{\tilde{\lambda}_1, \theta_0}(\theta)$ and $J_{\theta_0}(\theta)$ in a left neighborhood of θ_0 , namely when $\theta \in (\theta_0 - \delta, \theta_0)$ for $\delta > 0$ small enough. Therefore it suffices to compare $I''_{\tilde{\lambda}_1, \theta_0}(\theta_0)$ and the left second derivative $\frac{d^2}{d\theta^2} J_{\theta_0}(\theta-) \Big|_{\theta=\theta_0}$ which coincides with $J''_{\theta_0}(\theta_0)$ presented above for the

light-tailed case. We already explained that $\left. \frac{d^2}{d\theta^2} J_{\theta_0}(\theta-) \right|_{\theta=\theta_0}$ is an increasing function of $\rho \in (0, \infty)$; moreover

$$\left. \frac{d^2}{d\theta^2} J_{\theta_0}(\theta-) \right|_{\theta=\theta_0} = \frac{1}{\theta_0^2 \left[\frac{\Gamma(1+2/\rho)}{\Gamma^2(1+1/\rho)} - 1 \right]} \rightarrow 0 \text{ as } \rho \rightarrow 0$$

by taking into account the asymptotic behavior of Gamma function. In conclusion there exists $\rho_0 \simeq 0.81068$ (computed numerically) such that:

1. $I''_{\tilde{\lambda}_1, \theta_0}(\theta_0) > \left. \frac{d^2}{d\theta^2} J_{\theta_0}(\theta-) \right|_{\theta=\theta_0}$ for $\rho \in (0, \rho_0)$;
2. $I''_{\tilde{\lambda}_1, \theta_0}(\theta_0) < \left. \frac{d^2}{d\theta^2} J_{\theta_0}(\theta-) \right|_{\theta=\theta_0}$ for $\rho \in (\rho_0, 1)$.

Thus, by taking into account Remark 2.1, MQ estimators (with $\lambda = \tilde{\lambda}_1$) converge faster than MM estimators in the first case while, in the second case, the convergence of MQ and MM estimators cannot be compared because we cannot find $\delta > 0$ such that $I_{\tilde{\lambda}_1, \theta_0}(\theta) > J_{\theta_0}(\theta)$ or $I_{\tilde{\lambda}_1, \theta_0}(\theta) < J_{\theta_0}(\theta)$ for $0 < |\theta - \theta_0| < \delta$.

Remark 6.1 (The case $\rho = 1$, namely the exponential distribution). *If $\rho = 1$ the rate functions J_{θ_0} and $\Lambda_{\theta_0}^*$ coincide (in fact $\Gamma(1 + 1/\rho) = 1$) and an explicit expression of $\Lambda_{\theta_0}^*$ is available, namely*

$$J_{\theta_0}(\theta) = \Lambda_{\theta_0}^*(\theta) = \begin{cases} \frac{\theta}{\theta_0} - 1 - \log\left(\frac{\theta}{\theta_0}\right) & \text{for } \theta \in (0, \infty) \\ \infty & \text{otherwise;} \end{cases}$$

then we can directly compute $J''_{\theta_0}(\theta_0) = \frac{1}{\theta_0^2}$, which meets the above expression. In this case the MM estimators coincide with the ML estimators, and we already expected that they converge faster than the MQ estimators.

Analysis of MQ estimators for particular examples with $(\alpha, \omega) = (-\infty, \infty)$. Here we present the results concerning the specific examples listed above. For all cases except the one with the Gumbel distribution we choose η in order to have $G(0) \in \{0.25, 0.5, 0.75\}$, and we have some common features: $G(0) = 1/2$ for $\eta = 0$ (actually the symmetry property holds); we obtain symmetric values with respect $\lambda = 1/2$ such that the more the tails of the distributions are light, the more the numerical values of λ_{\max} are distant from $\lambda = 1/2$. The case with the Gumbel distribution behaves differently because the symmetry property fails for each fixed value of η . In all cases we can only give numerical values.

Normal distribution (namely G in (5)). We have $G(0) = \Phi(-\eta)$ and, by (9),

$$I''_{\lambda, \theta_0}(\theta_0) = \frac{(\varphi(\Phi^{-1}(\lambda))\{\eta + \Phi^{-1}(\lambda)\})^2}{\lambda(1-\lambda)\theta_0^2},$$

where φ is the standard Normal probability density function. Moreover

	numerical values for λ_{\max}
$\eta = 0$	0.06 (and 0.94 by symmetry)
$\eta = -\Phi^{-1}(1/4)$	0.90
$\eta = -\Phi^{-1}(3/4)$	$1 - 0.90 = 0.10$ (by symmetry)

Cauchy distribution (namely G in (6)). We have $G(0) = \frac{1}{\pi} (\arctan(-\eta) + \frac{\pi}{2})$ and, by (9),

$$I''_{\lambda, \theta_0}(\theta_0) = \frac{\{\eta + \tan((\lambda - \frac{1}{2})\pi)\}^2}{\pi^2 \{1 + \tan^2((\lambda - \frac{1}{2})\pi)\}^2 \lambda(1-\lambda)\theta_0^2}.$$

Moreover

	numerical values for λ_{\max}
$\eta = 0$	0.21 (and 0.79 by symmetry)
$\eta = 1$	0.65
$\eta = -1$	$1 - 0.65 = 0.35$ (by symmetry)

Logistic distribution (namely G in (7)). We have $G(0) = \frac{1}{1+e^\eta}$ and, by (9),

$$I''_{\lambda, \theta_0}(\theta_0) = \frac{\lambda(1-\lambda) \left(\eta - \log\left(\frac{1}{\lambda} - 1\right)\right)^2}{\theta_0^2}.$$

Moreover

	numerical values for λ_{\max}
$\eta = 0$	0.08 (and 0.92 by symmetry)
$\eta = \log 3$	0.85
$\eta = -\log 3$	$1 - 0.85 = 0.15$ (by symmetry)

Gumbel distribution (namely G in (8)). We have $G(0) = \exp(-e^\eta)$ and, by (9),

$$I''_{\lambda, \theta_0}(\theta_0) = \frac{\lambda(\log \lambda)^2 (\eta - \log(-\log \lambda))^2}{(1-\lambda)\theta_0^2}.$$

Some numerical inspections reveal that in general, for each fixed value of η , we can find an optimal value $\lambda_{\max} = \lambda_{\max}(\eta)$. Then, if we consider the value $\tilde{\lambda}_2$ and the function f_2 in Lemma 6.1(ii), we can say that

$$\lambda_{\max} = \lambda_{\max}(\eta) \rightarrow \tilde{\lambda}_2 \text{ as } |\eta| \rightarrow \infty$$

because $I''_{\lambda, \theta_0}(\theta_0)$ behaves like $\frac{f_2(\lambda)\eta^2}{\theta_0^2}$ when $|\eta|$ is large.

MM versus MQ estimators for particular examples with $(\alpha, \omega) = (-\infty, \infty)$. The MM estimators are well-defined only for the case with Gumbel distribution and, in the spirit of Remark 2.1, we can compare $J''_{\theta_0}(\theta_0)$ and $I''_{\lambda_{\max}, \theta_0}(\theta_0)$. However, for Normal and logistic distributions, it is possible to consider suitable GMM estimators $\{\tilde{\Theta}_n : n \geq 1\}$ by matching empirical and theoretical variances; so we present the rate function \tilde{J}_{θ_0} which governs the LDP of $\{\tilde{\Theta}_n : n \geq 1\}$ and, at least for the case of Normal distribution, we can give an expression of $\tilde{J}_{\theta_0}''(\theta_0)$ and we can compare the convergence of MQ and GMM estimators.

Gumbel distribution. We have $\mu(\theta) := \eta + \theta\gamma_*$, where γ_* is the Euler's constant, and therefore $\mu^{-1}(m) := \frac{m-\eta}{\gamma_*}$. Thus, by Proposition 4.1(i), $\{\mu^{-1}(\bar{X}_n) : n \geq 1\}$ satisfies the LDP with rate function J_{θ_0} with $c_1 = \frac{1}{\gamma_*}$ and $c_0 = \frac{-\eta}{\gamma_*}$. The rate functions J_{θ_0} and $\Lambda_{\theta_0}^*$ are good (we take into account Remark 4.1) and we can refer to the comparison between $I''_{\lambda_{\max}, \theta_0}(\theta_0)$ and $J''_{\theta_0}(\theta_0)$. We remark that, by Proposition 4.2, we have $J''_{\theta_0}(\theta_0) = \frac{6\gamma_*^2}{\theta_0^2\pi^2}$ for each fixed value of η (in fact we have $\sigma^2(\theta) = \frac{\theta^2\pi^2}{6}$). Then, for all $\eta \in \mathbb{R}$, we have $I''_{\lambda_{\max}, \theta_0}(\theta_0) > J''_{\theta_0}(\theta_0)$ noting that, for $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ as in Lemma 6.1, we have $\max\{I''_{\tilde{\lambda}_1, \theta_0}(\theta_0), I''_{\tilde{\lambda}_2, \theta_0}(\theta_0)\} > J''_{\theta_0}(\theta_0)$ (see Figure 1).

Normal (and logistic) distribution. The MM estimators are not well-defined because $\mu(\theta) := \eta$. So it is natural to match empirical and theoretical variances, i.e.

$$\sigma^2(\theta) = \frac{1}{n} \sum_{i=1}^n (x_i - \eta)^2, \text{ where } \sigma^2(\theta) = c\theta^2 \text{ and } c = \begin{cases} 1 & \text{for the Normal distribution} \\ \frac{\pi^2}{3} & \text{for the logistic distribution,} \end{cases}$$

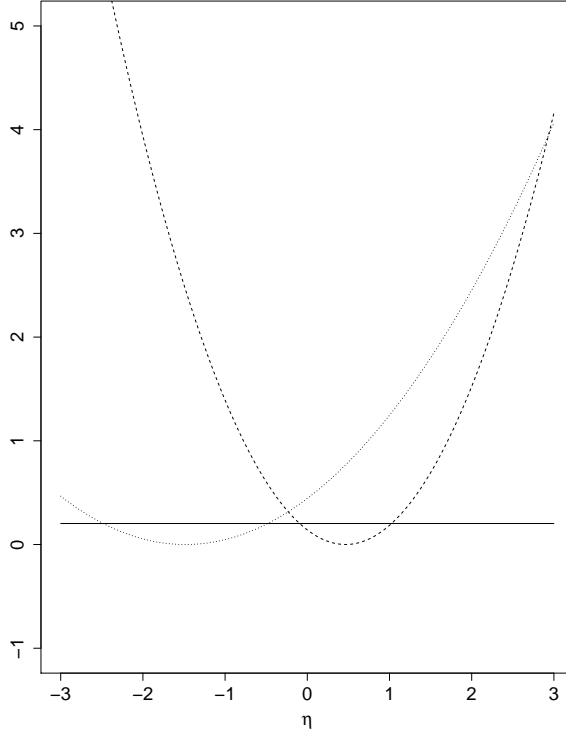


Figure 1: The second derivatives $I''_{\tilde{\lambda}_2, \theta_0}(\theta_0)$ (broken line) and $I''_{\tilde{\lambda}_1, \theta_0}(\theta_0)$ (dotted line) as functions of η . The solid line represents the value of $J''_{\theta_0}(\theta_0)$ which does not depend on η .

and we obtain the GMM estimators $\{\tilde{\Theta}_n : n \geq 1\}$ defined by

$$\tilde{\Theta}_n := \left(\frac{1}{cn} \sum_{i=1}^n (x_i - \eta)^2 \right)^{1/2}.$$

Then, by adapting the proof of Proposition 4.1, we can consider the function

$$\tilde{\Lambda}_{\theta_0}^*(y) := \sup_{\gamma \in \mathbb{R}} \left\{ \gamma y - \tilde{\Lambda}_{\theta_0}(\gamma) \right\}, \text{ where } \tilde{\Lambda}_{\theta_0}(\gamma) := \log \int_{\alpha_\theta}^{\omega_\theta} e^{\gamma(x-\eta)^2} dF_{\theta_0}(x),$$

and we can say $\{\tilde{\Theta}_n : n \geq 1\}$ satisfies the LDP with good rate function \tilde{J}_{θ_0} defined by

$$\tilde{J}_{\theta_0}(\theta) := \inf \{ \tilde{\Lambda}_{\theta_0}^*(y) : (y/c)^{1/2} = \theta \}.$$

From now on we restrict the attention to the case with Normal distribution because we can give explicit formulas. We have

$$\tilde{\Lambda}_{\theta_0}(\gamma) = \begin{cases} \frac{1}{2} \log \left(\frac{\theta_0^2/2}{\theta_0^2/2 - \gamma} \right) & \text{if } \gamma < \frac{\theta_0^2}{2} \\ \infty & \text{if } \gamma \geq \frac{\theta_0^2}{2}, \end{cases} \quad \tilde{\Lambda}_{\theta_0}^*(y) = \begin{cases} \frac{1}{2} \left[\frac{y}{\theta_0^2} - 1 - \log \left(\frac{y}{\theta_0^2} \right) \right] & \text{if } y > 0 \\ \infty & \text{if } y \leq 0, \end{cases}$$

and

$$\tilde{J}_{\theta_0}(\theta) = \begin{cases} \frac{1}{2} \left[\frac{\theta^2}{\theta_0^2} - 1 - \log \left(\frac{\theta^2}{\theta_0^2} \right) \right] & \text{if } \theta > 0 \\ \infty & \text{if } \theta \leq 0; \end{cases}$$

thus, after some computations, we get $\tilde{J}_{\theta_0}''(\theta_0) = \frac{2}{\theta_0^2}$ for all value of $\eta \in \mathbb{R}$. We conclude with the comparison between MQ and GMM estimators. Some numerical inspections reveal that in general, for each fixed value of η , we can find an optimal value $\lambda_{\max} = \lambda_{\max}(\eta)$ (their numerical values for $\eta = 0$, $\eta = \Phi^{-1}(1/4)$ and $\eta = \Phi^{-1}(3/4)$ were presented above); moreover $I_{\lambda_{\max}(\eta), \theta_0}''(\theta_0) > \tilde{J}_{\theta_0}''(\theta_0)$ for $|\eta|$ large enough because, for each fixed $\lambda \in (0, 1)$,

$$I_{\lambda, \theta_0}''(\theta_0) \rightarrow \infty \text{ as } |\eta| \rightarrow \infty.$$

On the other hand we cannot say that $I_{\lambda_{\max}(\eta), \theta_0}''(\theta_0) > \tilde{J}_{\theta_0}''(\theta_0)$ for all $\eta \in \mathbb{R}$; in fact, for $\eta = 0$, we have

$$I_{\lambda_{\max}(0), \theta_0}''(\theta_0) \simeq \frac{0.6085}{\theta_0^2} < \frac{2}{\theta_0^2} = \tilde{J}_{\theta_0}''(\theta_0)$$

(where $\lambda_{\max}(0) \simeq 0.06$ or $\lambda_{\max}(0) \simeq 0.94$). For completeness, following the same lines of the particular case with Gumbel distribution, we remark that

$$\lambda_{\max} = \lambda_{\max}(\eta) \rightarrow 1/2 \text{ as } |\eta| \rightarrow \infty$$

because, if we consider the function $f(\lambda) := \frac{(\varphi(\Phi^{-1}(\lambda)))^2}{\lambda(1-\lambda)}$, $I_{\lambda, \theta_0}''(\theta_0)$ behaves like $\frac{f(\lambda)\eta^2}{\theta_0^2}$ when $|\eta|$ is large, and $\sup_{\lambda \in (0,1)} f(\lambda) = f(1/2)$.

6.2 Example 2

We start with the analysis of MQ estimators. We conclude with the MM estimators, and their comparison with the MQ estimators.

Analysis of MQ estimators. Here we present the results concerning the specific examples listed above. In all cases, except the one with Gumbel distribution, we can conclude that $\lambda = 1/2$ is optimal; however we can find counterexamples (see Appendix A). A further common feature (for all cases except the one with Gumbel distribution) is that $\left. \frac{d}{d\lambda} I_{\lambda, \theta_0}''(\theta_0) \right|_{\lambda=1/2} = 0$ (and obviously this does not guarantee that $\lambda = 1/2$ is an optimal; this will be explained in Appendix A); in fact we have

$$\begin{aligned} \left. \frac{d}{d\lambda} I_{\lambda, \theta_0}''(\theta_0) \right|_{\lambda=1/2} &= \left. \frac{2G'(G^{-1}(\lambda))G''(G^{-1}(\lambda))(G^{-1})'(\lambda)\lambda(1-\lambda) - (1-2\lambda)\{G'(G^{-1}(\lambda))\}^2}{\lambda^2(1-\lambda)^2} \right|_{\lambda=1/2} \\ &= \left. \frac{2G''(G^{-1}(\lambda))\lambda(1-\lambda) - (1-2\lambda)\{G'(G^{-1}(\lambda))\}^2}{\lambda^2(1-\lambda)^2} \right|_{\lambda=1/2} = 0 \end{aligned}$$

noting that $G^{-1}(1/2) = 0$ (because the distribution function G has the symmetry property) and $G''(0) = 0$ (because the probability density function $G'(x)$ has a maximum at $x = 0$).

Normal distribution (namely G in (10)). By (14) we have

$$I_{\lambda, \theta_0}''(\theta_0) = \frac{\varphi^2(\Phi^{-1}(\lambda))}{s^2\lambda(1-\lambda)}.$$

One can check numerically that we have a unique optimal λ_{\max} (for every s), namely $\lambda_{\max} = 0.5$. Cauchy distribution (namely G in (11)). By (14) we have

$$I_{\lambda, \theta_0}''(\theta_0) = \frac{1}{s^2\pi^2 \left\{1 + \tan^2\left((\lambda - \frac{1}{2})\pi\right)\right\}^2 \lambda(1-\lambda)}.$$

One can check numerically that we have a unique optimal λ_{\max} (for every s), namely $\lambda_{\max} = 0.5$.
Logistic distribution (namely G in (12)). By (14) we have

$$I''_{\lambda, \theta_0}(\theta_0) = \frac{\lambda(1 - \lambda)}{s^2}$$

One can immediately check (we have a polynomial with degree 2) that we have a unique optimal λ_{\max} (for every s), namely $\lambda_{\max} = 0.5$.

Gumbel distribution (namely G in (13)). By (14) we have

$$I''_{\lambda, \theta_0}(\theta_0) = \frac{\lambda(\log \lambda)^2}{s^2(1 - \lambda)}$$

Then we have a unique optimal value $\lambda_{\max} = \tilde{\lambda}_2$ (for every s), where $\tilde{\lambda}_2$ is the value in Lemma 6.1(ii); in fact, if we consider the function f_2 in that lemma, we have $I''_{\lambda, \theta_0}(\theta_0) = \frac{f_2(\lambda)}{s^2}$.

MM versus MQ estimators. The MM estimators are well-defined in all cases except the one with Cauchy distribution. Moreover, by taking into account Remark 4.1, we can always refer to the comparison between $J''_{\theta_0}(\theta_0)$ and $I''_{\lambda_{\max}, \theta_0}(\theta_0)$.

Normal distribution. In this case $\mu(\theta) := \theta$. Thus, by Proposition 4.1(i), $\{\mu^{-1}(\bar{X}_n) : n \geq 1\} = \{\bar{X}_n : n \geq 1\}$ satisfies the LDP with rate function J_{θ_0} defined by

$$J_{\theta_0}(\theta) = \Lambda_{\theta_0}^*(\theta) = \frac{(\theta - \theta_0)^2}{2s^2}$$

(J_{θ_0} and $\Lambda_{\theta_0}^*$ coincide noting that $c_1 = 1$ and $c_0 = 0$). Then, since $\lambda_{\max} = 1/2$ we have to compare $J''_{\theta_0}(\theta_0) = \frac{1}{s^2}$ (which meets the expression provided by Proposition 4.2 noting that $\sigma^2(\theta) = s^2$) and $I''_{1/2, \theta_0}(\theta_0) = \frac{2}{\pi s^2}$ and, obviously, we have $J''_{\theta_0}(\theta_0) > I''_{1/2, \theta_0}(\theta_0)$ (for every s). Thus MM estimators converge faster than every MQ estimators; in some sense we already expected this noting that the MM estimators coincide with the ML estimators.

Logistic distribution. In this case $\mu(\theta) := \theta$. Thus, by Proposition 4.1(i), $\{\mu^{-1}(\bar{X}_n) : n \geq 1\} = \{\bar{X}_n : n \geq 1\}$ satisfies the LDP with rate function J_{θ_0} , which coincides with $\Lambda_{\theta_0}^*$ (we have again $c_1 = 1$ and $c_0 = 0$); in this case we cannot provide an explicit expression of $\Lambda_{\theta_0}^*$. Then, since $\lambda_{\max} = 1/2$ we have to compare $J''_{\theta_0}(\theta_0) = \frac{3}{\pi^2 s^2}$ (this is a consequence of Proposition 4.2 noting that $\sigma^2(\theta) = \frac{\pi^2 s^2}{3}$) and $I''_{1/2, \theta_0}(\theta_0) = \frac{1}{4s^2}$ and, obviously, we have $J''_{\theta_0}(\theta_0) > I''_{1/2, \theta_0}(\theta_0)$ (for every s). Thus MM estimators converge faster than every MQ estimators but, differently from what happens for the case with Normal distribution, they do not coincide with ML estimators.

Gumbel distribution. In this case $\mu(\theta) := \theta + s\gamma_*$, where γ_* is the Euler's constant. Thus, by Proposition 4.1(i), $\{\mu^{-1}(\bar{X}_n) : n \geq 1\} = \{\bar{X}_n : n \geq 1\}$ satisfies the LDP with rate function J_{θ_0} (with $c_1 = 1$ and $c_0 = -s\gamma_*$); in this case we cannot provide an explicit expression of $\Lambda_{\theta_0}^*$. The rate functions J_{θ_0} and $\Lambda_{\theta_0}^*$ are good and we can refer to the comparison between

$$J''_{\theta_0}(\theta_0) = \frac{6}{\pi^2 s^2}$$

(this value is a consequence of Proposition 4.2 noting that $\sigma^2(\theta) = \frac{\pi^2 s^2}{6}$) and, for the optimal value $\lambda_{\max} = \tilde{\lambda}_2$ defined in Lemma 6.1(ii),

$$I''_{\tilde{\lambda}_2, \theta_0}(\theta_0) = \frac{\tilde{\lambda}_2(\log \tilde{\lambda}_2)^2}{s^2(1 - \tilde{\lambda}_2)} = \frac{\tilde{\lambda}_2(2\tilde{\lambda}_2 - 2)^2}{s^2(1 - \tilde{\lambda}_2)} = \frac{4\tilde{\lambda}_2(1 - \tilde{\lambda}_2)}{s^2}.$$

We can check numerically that $I''_{\tilde{\lambda}_2, \theta_0}(\theta_0) > J''_{\theta_0}(\theta_0)$ (for every s); in fact we have $4\tilde{\lambda}_2(1-\tilde{\lambda}_2) \simeq 0.6476$ (we get a numerical value obtained for the statistical model with Weibull distributions because $4\tilde{\lambda}_2(1-\tilde{\lambda}_2) = 4\tilde{\lambda}_1(1-\tilde{\lambda}_1)$ by Lemma 6.1(iii)) and $\frac{6}{\pi^2} \simeq 0.6079$. Thus MQ estimators with the optimal value λ_{\max} converge faster than MM estimators.

6.3 Example 3

Here we analyze the MQ estimators for the specific examples listed above. In all cases we can only give numerical values; such values depend on the unknown parameter θ_0 , and therefore we do not discuss the comparison with the MM estimators (as we do for the other examples). We have the same feature highlighted for Example 1 with $(\alpha, \omega) = (-\infty, \infty)$, namely the more the tail of the distributions are light, the more the numerical values of λ_{\max} are distant from $\lambda = 1/2$.

Normal distribution (namely G in (10)). By (15) we have

$$I''_{\lambda, \theta_0}(\theta_0) = \begin{cases} \frac{1}{\lambda(1-\lambda)} \left[\varphi(\Phi^{-1}(\frac{\lambda}{1+\theta_0}))\Phi^{-1}(\frac{\lambda}{1+\theta_0}) - \frac{\lambda}{1+\theta_0} \right]^2 & \text{for } \lambda \in (0, \frac{1+\theta_0}{2}] \\ \frac{1}{\lambda(1-\lambda)} \left[-\varphi(\Phi^{-1}(\frac{\lambda-\theta_0}{1-\theta_0}))\Phi^{-1}(\frac{\lambda-\theta_0}{1-\theta_0}) + \frac{\lambda-1}{1-\theta_0} \right]^2 & \text{for } \lambda \in (\frac{1+\theta_0}{2}, 1). \end{cases}$$

Moreover

	numerical values for λ_{\max}
$\theta_0 = 0$	0.15 (and 0.85 by symmetry)
$\theta_0 = 1/2$	0.94
$\theta_0 = -1/2$	$1 - 0.94 = 0.06$ (by symmetry)

Cauchy distribution (namely G in (11)). By (15) we have

$$I''_{\lambda, \theta_0}(\theta_0) = \begin{cases} \frac{1}{\lambda(1-\lambda)} \left[\frac{\tan((\frac{\lambda}{1+\theta_0} - \frac{1}{2})\pi)}{\pi(1+\tan^2((\frac{\lambda}{1+\theta_0} - \frac{1}{2})\pi))} - \frac{\lambda}{1+\theta_0} \right]^2 & \text{for } \lambda \in (0, \frac{1+\theta_0}{2}] \\ \frac{1}{\lambda(1-\lambda)} \left[-\frac{\tan((\frac{\lambda-\theta_0}{1-\theta_0} - \frac{1}{2})\pi)}{\pi(1+\tan^2((\frac{\lambda-\theta_0}{1-\theta_0} - \frac{1}{2})\pi))} + \frac{\lambda-1}{1-\theta_0} \right]^2 & \text{for } \lambda \in (\frac{1+\theta_0}{2}, 1). \end{cases}$$

Moreover

	numerical values for λ_{\max}
$\theta_0 = 0$	0.39 (and 0.61 by symmetry)
$\theta_0 = 1/2$	0.84
$\theta_0 = -1/2$	$1 - 0.84 = 0.16$ (by symmetry)

Logistic distribution (namely G in (12)). By (15) we have

$$I''_{\lambda, \theta_0}(\theta_0) = \begin{cases} \frac{1}{\lambda(1-\lambda)} \left[-\frac{\lambda}{1+\theta_0} (1 - \frac{\lambda}{1+\theta_0}) \log(\frac{1+\theta_0}{\lambda} - 1) - \frac{\lambda}{1+\theta_0} \right]^2 & \text{for } \lambda \in (0, \frac{1+\theta_0}{2}] \\ \frac{1}{\lambda(1-\lambda)} \left[\frac{\lambda-\theta_0}{1-\theta_0} (1 - \frac{\lambda-\theta_0}{1-\theta_0}) \log(\frac{1-\theta_0}{\lambda-\theta_0} - 1) + \frac{\lambda-1}{1-\theta_0} \right]^2 & \text{for } \lambda \in (\frac{1+\theta_0}{2}, 1). \end{cases}$$

Moreover

	numerical values for λ_{\max}
$\theta_0 = 0$	0.22 (and 0.78 by symmetry)
$\theta_0 = 1/2$	0.92
$\theta_0 = -1/2$	$1 - 0.92 = 0.08$ (by symmetry)

6.4 Example 4

Here we analyze Example 4. For MQ estimators we have the same rate function presented in Example 1 with $(\alpha, \omega) = (0, \infty)$ when G is as in (4) and $\rho = 1$. Thus we have a unique optimal λ_{\max} which does not depend on θ_0 , namely $\lambda_{\max} = \tilde{\lambda}_1$ where $\tilde{\lambda}_1$ is defined in Lemma 6.1(i).

Now we briefly discuss the MM estimators for Example 4. We recall that $\mu(\theta)$ is finite only if $\theta \in \tilde{\Theta} := (0, 1)$, where $\tilde{\Theta} \subset \Theta = (0, \infty)$. So we could consider the mean function on the restricted parameter space $\tilde{\Theta}$, i.e.

$$\mu(\theta) = \frac{1/\theta}{1/\theta - 1} = \frac{1}{1 - \theta} \text{ for } \theta \in \tilde{\Theta}.$$

Then, if we consider the restricted parameter space $\tilde{\Theta}$, the MM estimators $\{\mu^{-1}(\bar{X}_n) : n \geq 1\}$ are defined by $\mu^{-1}(\bar{X}_n) = 1 - \bar{X}_n^{-1}$, and the function $\mu^{-1}(\cdot)$ is continuous on $(\alpha, \omega) = (1, \infty)$. Unfortunately we cannot apply Proposition 4.1 because we cannot consider neither the hypotheses of Proposition 4.1(i) (obvious) nor the hypotheses of Proposition 4.1(ii) because Pareto distributions are heavy-tailed and $\Lambda_{\theta_0}^*$ is not good (see Remark 4.1).

6.5 Example 5

Here we analyze Example 5. As far as the MQ estimators are concerned, we can say that we cannot find an optimal λ because $I''_{\lambda, \theta_0}(\theta_0)$ is an increasing function; in fact, by (17), the derivative of $I''_{\lambda, \theta_0}(\theta_0)$ with respect to λ is

$$\frac{d}{d\lambda} I''_{\lambda, \theta_0}(\theta_0) = \frac{(G'(\theta_0))^2}{(1 - \lambda)^2 G^2(\theta_0)}.$$

We can also say that the larger is λ the faster is the convergence of the MQ estimators.

In the remaining part we deal with as the MM estimators, and their comparison with the MQ estimators. Obviously the rate functions J_{θ_0} and $\Lambda_{\theta_0}^*$ are good (we take into account Remark 4.1) and we can refer to the comparison between $J''_{\theta_0}(\theta_0)$ and $I''_{\lambda, \theta_0}(\theta_0)$ (for $\lambda \in (0, 1)$); for completeness we remark that we cannot obtain an explicit expression of $\Lambda_{\theta_0}^*$ (even for the simplest case with the uniform distributions, i.e. the case $G(x) = x$ for all $x \in (0, \infty)$). It is easy to check that, if we consider λ_0 defined by

$$\lambda_0 := \frac{(\mu'(\theta_0))^2 G^2(\theta_0)}{(\mu'(\theta_0))^2 G^2(\theta_0) + \sigma^2(\theta_0) (G'(\theta_0))^2}, \quad (18)$$

we have

$$J''_{\theta_0}(\theta_0) = \frac{(\mu'(\theta_0))^2}{\sigma^2(\theta_0)} > \frac{\lambda (G'(\theta_0))^2}{(1 - \lambda) G^2(\theta_0)} = I''_{\lambda, \theta_0}(\theta_0) \text{ for } \lambda < \lambda_0$$

and

$$J''_{\theta_0}(\theta_0) = \frac{(\mu'(\theta_0))^2}{\sigma^2(\theta_0)} < \frac{\lambda (G'(\theta_0))^2}{(1 - \lambda) G^2(\theta_0)} = I''_{\lambda, \theta_0}(\theta_0) \text{ for } \lambda > \lambda_0$$

by Proposition 4.2 and (17).

Thus the MQ estimators converge faster than MM estimators if λ is close to 1; this is not surprising because the case $\lambda = 1$ concerns the case of ML estimators $\{X_{n:n} : n \geq 1\}$. For completeness we can say that, if $\{X_n : n \geq 1\}$ are i.i.d. with distribution function F_{θ_0} as in Example 5, the LDP in Proposition 2.1 with $\lambda = 1$ is governed by a good rate function; thus we can consider a version of Proposition 3.1 with $\lambda = 1$, and we have the LDP of $\{X_{n:n} : n \geq 1\}$ with good rate function I_{1, θ_0} defined by

$$I_{1, \theta_0}(\theta) := \begin{cases} \log \frac{1}{h_{\lambda, \theta_0}(\theta)} = \log \frac{G(\theta_0)}{G(\theta)} & \text{for } \theta \in \Theta \text{ such that } \theta \in (0, \theta_0) \\ \infty & \text{otherwise.} \end{cases}$$

We also remark that in general the threshold value λ_0 in (18) depends on θ_0 . In fact, for $G(x) := e^x - 1$, after some computations we have $\mu(\theta) = \frac{\theta e^\theta}{e^\theta - 1} - 1$, $\sigma^2(\theta) = \frac{(e^\theta - 1)^2 - \theta^2 e^\theta}{(e^\theta - 1)^2}$, and therefore

$$\lambda_0 = \frac{e^{2\theta_0} - 2e^{\theta_0}(1 + \theta_0) + (1 + \theta_0)^2}{2e^{2\theta_0} - e^{\theta_0}(4 + 2\theta_0 + \theta_0^2) + 2 + 2\theta_0 + \theta_0^2}.$$

Interestingly we can say that λ_0 does not depend on θ_0 if $G(x) := x^y$ for some $y > 0$; in fact we have $\mu(\theta) = \frac{y\theta}{y+1}$, $\sigma^2(\theta) = \frac{y\theta^2}{(y+2)(y+1)^2}$, and therefore

$$\lambda_0 = \frac{(\frac{y}{y+1})^2(\theta_0^y)^2}{(\frac{y}{y+1})^2(\theta_0^y)^2 + \frac{y\theta_0^2}{(y+2)(y+1)^2}(y\theta_0^{y-1})^2} = \frac{1}{1 + \frac{y}{y+2}} = \frac{y+2}{2y+2}.$$

For instance, for the specific case of uniform distributions cited in Example 5 (for which we have $\mu(\theta) = \frac{\theta}{2}$ and $\sigma^2(\theta) = \frac{\theta^2}{12}$ for all $\theta \in (0, \infty)$; so the sequence $\{\mu^{-1}(\bar{X}_n) : n \geq 1\}$ in Proposition 4.1 is defined by $\mu^{-1}(\bar{X}_n) = 2\bar{X}_n$) we have $G(x) := x$, and therefore we get $\lambda_0 = 3/4$ by setting $y = 1$.

Finally we remark that, in general, we cannot find $\delta > 0$ such that $I_{\lambda_0, \theta_0}(\theta) > J_{\theta_0}(\theta)$ or $I_{\lambda_0, \theta_0}(\theta) < J_{\theta_0}(\theta)$ for $0 < |\theta - \theta_0| < \delta$; for instance (see Figure 2 where $\theta_0 = 1$) this happens for the statistical model with uniform distributions cited above (where $G(x) := x$ and $\lambda_0 = 3/4$).

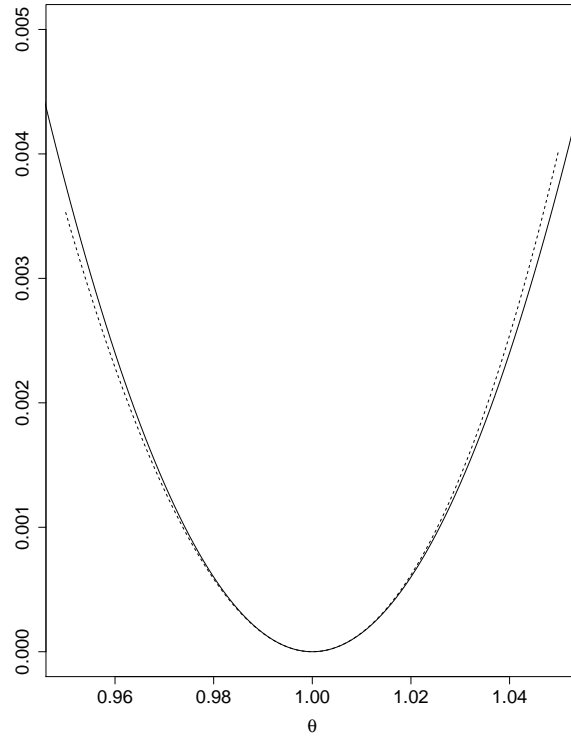


Figure 2: The rate functions $I_{3/4, \theta_0}(\theta)$ (broken line) and $J_{\theta_0}(\theta)$ (solid line) in a neighborhood of $\theta_0 = 1$ for the statistical model with uniform distributions.

A A class of counterexamples

In Section 6.2, for all the examples where the distribution G is symmetric, we find that $G''(0) = 0$ and that there is a unique optimal value λ_{\max} , namely $\lambda_{\max} = 0.5$.

Here we show that this is not necessarily the case, indeed we present a procedure to construct another function \tilde{G} with the symmetry property and such that $\tilde{G}''(0) = 0$; this function will be determined starting from a function G with the properties cited above (for instance it could be one of the choices illustrated in Example 2 except the one concerning Gumbel distribution). The aim is to illustrate that, for such a function \tilde{G} , $\lambda = 0.5$ cannot be an optimal value.

The function \tilde{G} is defined by

$$\tilde{G}(x) := \begin{cases} \frac{G(1+x)}{2G'(0)+1} & \text{for } x \leq -1 \\ \frac{1}{2} + \frac{G'(0)x}{2G'(0)+1} & \text{for } |x| < 1 \\ \frac{2G'(0)+G(x-1)}{2G'(0)+1} & \text{for } x \geq 1. \end{cases}$$

One can check that, if G is twice differentiable, then \tilde{G} is also twice differentiable (and in particular the condition $G''(0) = 0$ is needed to say that \tilde{G} is twice differentiable); the details are omitted. Moreover, if we consider

$$G(-1) = \frac{1}{2(2G'(0)+1)} \text{ and } G(1) = \frac{4G'(0)+1}{2(2G'(0)+1)}$$

(we recall that $G(0) = \frac{1}{2}$ by the symmetry property of G), we have

$$\tilde{G}^{-1}(\lambda) := \begin{cases} G^{-1}(\lambda(2G'(0)+1)) - 1 & \text{for } \lambda \leq G(-1) \\ (\lambda - \frac{1}{2}) \frac{2G'(0)+1}{G'(0)} & \text{for } G(-1) < \lambda < G(1) \\ G^{-1}(\lambda(2G'(0)+1) - 2G'(0)) + 1 & \text{for } \lambda \geq G(1). \end{cases}$$

Then, around $\lambda = 1/2$ (more precisely for $\lambda \in (G(-1), G(1))$ because $G(-1) \in (0, \frac{1}{2})$ and $G(1) \in (\frac{1}{2}, 1)$), we have $I''_{\lambda, \theta_0}(\theta_0) = \frac{1}{\lambda(1-\lambda)} \left(\frac{G'(0)}{2G'(0)+1} \right)^2$ by (14); thus $\lambda = 0.5$ cannot be an optimal value because $I''_{\lambda, \theta_0}(\theta_0)$ is locally minimized at $\lambda = 1/2$ (in fact $\lambda = 1/2$ maximizes the denominator $\lambda(1-\lambda)$).

References

- [1] Aitchison J., Brown J.A.C. (1957) The Lognormal Distribution. Cambridge University Press, Cambridge.
- [2] Castillo E., Hadi A.S. (1995) A method for estimating parameters and quantiles of distributions of continuous random variables. Comput. Statist. Data Anal. 20, 421–439.
- [3] DasGupta A. (2008) Asymptotic Theory of Statistics and Probability. Springer, New York.
- [4] Dembo A., Zeitouni O. (1998) Large Deviations Techniques and Applications. 2nd Edition. Springer, New York.
- [5] Denuit M., Dhaene J., Goovaerts M., Kaas R. (2005) Actuarial Theory for Dependent Risks. John Wiley and Sons, Chichester.

- [6] Dominicy Y., Veredas D. (2013) The method of simulated quantiles. *J. Econometrics* 172, 235–247.
- [7] Hashorva E., Macci C., Pacchiarotti C. (2013) Large deviations for proportions of observations which fall in random sets determined by order statistics. *Methodol. Comput. Appl. Probab.* 15, 875–896.
- [8] Hassanein K.M. (1969a) Estimation of the parameters of the extreme value distribution by use of two or three order statistics. *Biometrika* 56, 429–436.
- [9] Hassanein K.M. (1969b) Estimation of the parameters of the logistic distribution by sample quantiles. *Biometrika* 56, 684–687.
- [10] Hassanein K.M. (1971) Percentile estimators for the parameters of the Weibull distribution. *Biometrika* 58, 673–676.
- [11] Hassanein K.M. (1972) Simultaneous estimation of the parameters of the extreme value distribution of sample quantiles. *Technometrics* 14, 63–70.
- [12] Koenker R. (2005) *Quantile regression*. Cambridge university press, Cambridge.
- [13] McNeil, A.J., Frey, R. and Embrechts, P. (2015) *Quantitative risk management: Concepts, techniques and tools*. Princeton university press, Princeton.
- [14] Mudholkar G.S., Hutson A.D. (2000) The epsilon-skew-normal distribution for analyzing near-normal data. *J. Statist. Plann. Inference* 83, 291–309.
- [15] Sgouropoulos N., Yao Q., Yastremiz C. (2015) Matching a distribution by matching quantiles estimation. *J. Amer. Statist. Assoc.* 110:510, 742–759.